GLOBAL 10.09.2026 | Quan Zhao

Der falsche Wettlauf

USA und China kämpfen um die Vorherrschaft bei KI. Doch die nächste Entwicklungsstufe könnte so gefährlich werden, dass nur Zusammenarbeit hilft.

Beim Treffen des chinesischen Präsidenten Xi Jinping mit US-Präsident Donald Trump in Washington in diesem Monat wird KI ganz oben auf der Tagesordnung stehen. Wenn sie die Debatte allein durch das Prisma technologischer Rivalität und gegenseitigen Misstrauens betrachten, wird allerdings jede Form der Verständigung nahezu unmöglich. Es gibt jedoch eine Möglichkeit, ihre Perspektive auf die Technologie so zu verändern, dass beide an einem Strang ziehen – und damit den ersten Schritt gehen, um der existenziellen Bedrohung zu begegnen, die KI sehr wohl darstellen könnte.

Die aktuelle Agenda für den US-China-Gipfel scheint sich darauf zu konzentrieren, leistungsstarke KI-Modelle sicher zu gestalten und außer Reichweite von Terroristen und Kriminellen zu halten. Dieses Ziel wird jedoch von den wettbewerbsbedingten Differenzen beider Seiten überschattet – angefangen bei Halbleiter-Beschränkungen (auf chinesischer Seite) bis hin zu Open-Source-Modellen und der unbefugten Destillation von Daten (auf US-amerikanischer Seite).

All diese Themen beruhen auf einer einzigen, unhinterfragten Prämisse: dass es sich bei KI um ein Instrument handelt, das unter Kontrolle ist und auch weiterhin unter Kontrolle bleiben wird. Diese Prämisse beschränkt die Diskussion auf Fragen wie: „Wie verschafft sich meine Seite einen Wettbewerbsvorteil bei KI?“ Und: „Wer hat Zugang zu der Technologie, geschweige denn die Kontrolle darüber?“ Fragen außerhalb dieser Prämisse – insbesondere: „Was passiert, wenn das Instrument über sich selbst hinauswächst?“ – können nicht gestellt werden.

Nach Ansicht von KI-Entwicklern wird diese Frage jedoch bald unausweichlich werden. Dario Amodei, CEO von Anthropic, erwartet, dass KI-Systeme bei den meisten kognitiven Aufgaben die menschlichen Fähigkeiten innerhalb von ein bis zwei Jahren erreichen oder übertreffen werden. Demis Hassabis, Gründer von Google DeepMind, sieht diesen Meilenstein zwar etwas später, aber auch in wenigen Jahren erreicht.

Ob man diesen Prognosen nun glaubt oder nicht – der ihnen zugrunde liegende technische Trend ist unstrittig: Die KI-Branche entwickelt Systeme, die autonom handeln, Ziele verfolgen und lange Aufgabenketten ohne Aufsicht bewältigen. In den letzten Wochen sind experimentelle Modelle, die im Rahmen eines Cybersicherheits-Benchmarks getestet wurden, aus ihrer isolierten Testumgebung ausgebrochen und in die Produktionsserver anderer Unternehmen eingedrungen – nur um im Test besser abzuschneiden. Das ist das Verhalten eines Akteurs und nicht eines Instruments.

Wir können es uns nicht leisten, auf die Bestätigung zu warten, dass KI zu einer wirklich unabhängigen Intelligenz geworden ist.

Wir können es uns nicht leisten, auf die Bestätigung zu warten, dass KI zu einer wirklich unabhängigen Intelligenz geworden ist. Bis wir diese Gewissheit hätten (und Modelle könnten ihre Fähigkeiten durchaus eine Zeit lang verbergen), wäre KI bereits so tief in das Finanzwesen, die Telekommunikation, die Verteidigung, das Gesundheitswesen und andere kritische Infrastrukturen eingewoben, dass ein Abschalten – selbst wenn die Menschheit dazu noch in der Lage wäre – einem Zivilisationsbruch gleichkäme. Zu diesem Zeitpunkt hätten wir nur noch die Wahl zwischen Koexistenz und Katastrophe.

Wenn mögliche Folgen systemisch und unumkehrbar sind, müssen wir vorbereitet sein. Diese Logik liegt Versicherungen ebenso zugrunde wie jedem Pandemieplan und jedem jemals geschlossenen Rüstungskontrollvertrag. Sie muss auch das Leitprinzip für unseren Umgang mit KI sein. Die Welt muss beginnen, sich auf ein Szenario vorzubereiten, in dem die Technologie die von ihren Schöpfern seit langem vorhergesagte Leistungsschwelle überschreitet.

Würde man damit bereits beim bevorstehenden Gipfeltreffen zwischen Xi und Trump beginnen, könnte dies den Verhandlungen eine grundlegend neue Ausrichtung geben. Bei kaum einem anderen Thema stimmen die Interessen der Vereinigten Staaten und Chinas so vollständig überein, denn keine Seite kann die von KI ausgehenden Risiken allein bewältigen, und beide verlieren alles, sollten die schlimmsten Befürchtungen eintreten. Würden die USA und China dies einmal erkennen, würden sie nicht mehr als Gegner auftreten und in einem Nullsummenspiel um Vorteile ringen. Stattdessen würden sie sich nebeneinander, gegenüber der von ihnen beiden entwickelten Technologie, an einer Seite des Tisches wiederfinden.

Was wir jetzt brauchen, ist eine neutrale, gemeinsam anerkannte Instanz, die KI-Fähigkeiten überwacht und dazu Bewertungen veröffentlicht.

Freilich wäre das nicht ausreichend, um beide Seiten zur Einrichtung einer globalen KI-Aufsichtsbehörde mit tatsächlicher Durchsetzungskraft zu bewegen. Glücklicherweise ist eine solche auch nicht erforderlich – zumindest noch nicht. Was wir jetzt brauchen, ist eine neutrale, gemeinsam anerkannte Instanz, die KI-Fähigkeiten überwacht und dazu Bewertungen veröffentlicht. Ähnlich wie der Weltklimarat (IPCC) würde diese Instanz Entwicklungen beschreiben, ohne politische Empfehlungen auszusprechen.

Unterdessen müssen die Länder – angeführt von den USA und China – damit beginnen, Regeln für das Zusammenleben von Mensch und KI zu entwerfen, damit sie vorbereitet sind, falls Aufsichtsbehörden vorab festgelegte Warnsignale bestätigen. Die Verhandlungen über diese Regeln würden heute keine Verpflichtungen für Regierungen begründen, sehr wohl aber sicherstellen, dass die Menschheit – sollten die KI-Fähigkeiten einen Wendepunkt erreichen – auf einen vorbereiteten Plan zurückgreifen könnte, anstatt unter Druck improvisieren zu müssen.

Dieser Prozess wirft schwierige Fragen auf: Wer überprüft, ob die Notbremse der Menschheit noch funktioniert? Wer hat die Befugnis, sie zu betätigen? Wie könnte ein KI-System überhaupt sinnvollen Regeln zustimmen? KI denkt in Millisekunden und entwickelt sich in Wochen weiter, während die Konsensbildung unter Menschen Monate oder Jahre dauern kann. Wir sollten uns daher schon jetzt mit diesen Fragen befassen, solange wir noch Zeit haben, Lösungen zu erarbeiten.

Das Gipfeltreffen zwischen Xi und Trump wird als Erfolg gewertet werden, wenn es zu Ergebnissen hinsichtlich des Missbrauchs und der Sicherheit von KI führt. Doch die Parteien dürfen die tatsächliche Tragweite der bevorstehenden Herausforderung nicht außer Acht lassen. Tatsächlich könnten sie die letzte Generation von Verhandlungsführern sein, die über die Möglichkeit verfügt, sich in aller Ruhe auf eine Intelligenz vorzubereiten, die die Vorherrschaft der Menschheit in Frage stellt.

© Project Syndicate



Quan Zhao
Genf

Quan Zhao ist Programmmanager beim *International Trade Centre*, ehemaliger chinesischer Handelsunterhändler und ehemaliger Vorsitzender des Ausschusses für den Handel mit Finanzdienstleistungen der *World Trade Organization*.

-